AJAY KADIYALA

Introduction – Master the

Data Engineering Roadmap

- **Title**: "Master the Data Engineering Roadmap"

- **Subtitle**: "Learn strategies and tools to excel in your Data Engineering career."

- **Presenter**: Ajay Kadiyala, Data Engineer

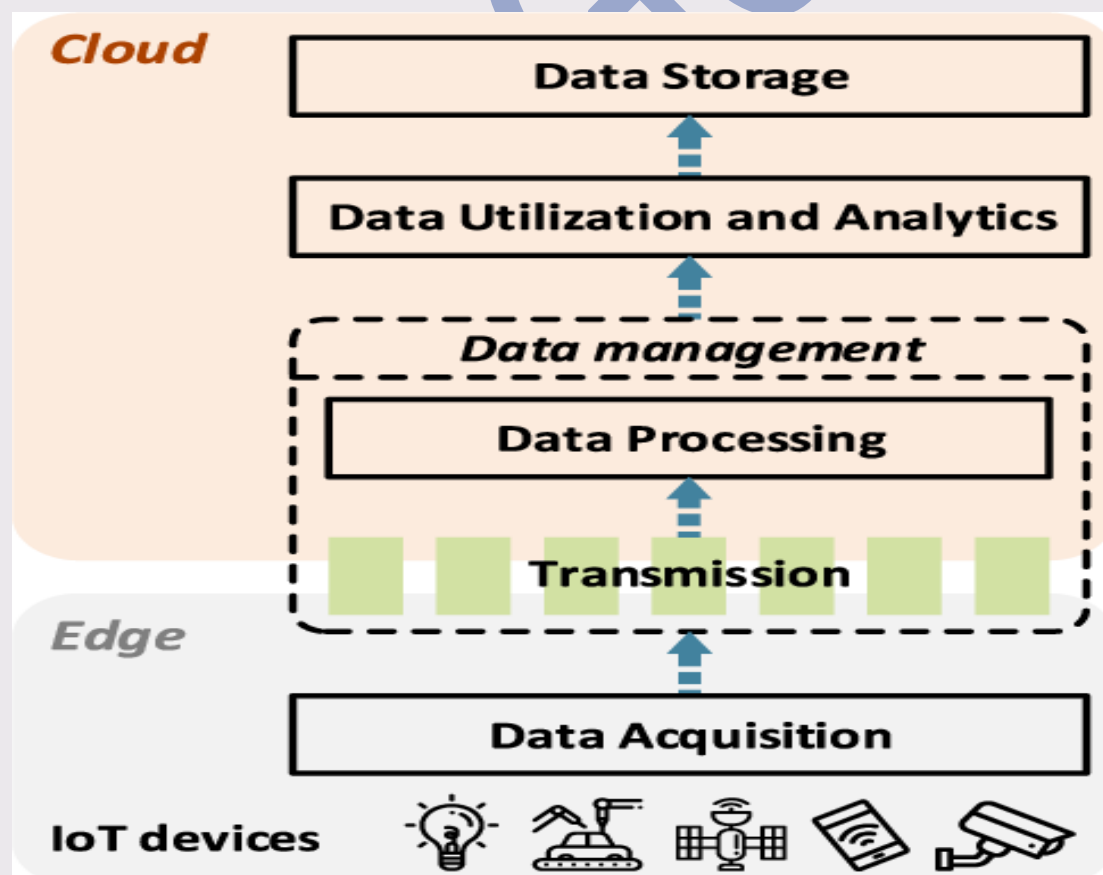- **Date**: 26th Oct 2024

Follow me Here:

LinkedIn: https://www.linkedin.com/in/ajay026/

Data Geeks Community: https://lnkd.in/gU5NkCqi

What is Data Engineering? And why is that important?

Data Engineering is the process of designing, building, and maintaining systems that enable companies to collect, store, and analyse data. It involves creating **data pipelines**, managing databases, and ensuring data flows smoothly and efficiently across systems.

Why it's important: In today's data-driven world, businesses rely on data to make decisions, **optimize** operations, and improve products. Data Engineers make it possible by organizing raw data into usable, reliable information, enabling teams to gain insights and drive growth.

- ➢ **Data Acquisition**: IoT devices capture raw data from their surroundings.
- ➢ **Transmission**: This data is sent to the next stage.
- ➢ **Data Processing**: The data is cleaned and processed to make it useful.
- ➢ **Data Management**: Organizes and manages processed data for efficient use.
- ➢ **Data Utilization and Analytics**: In the cloud, data is analysed for insights.
- ➢ **Data Storage**: Finally, data is stored in the cloud for future use or reference.

What Skills Required for you as a Data Engineer?

Step -1(Programming & Database)

- ✓ SQL
- ✓ Python
- ✓ Pyspark

Step -2 (Fundamentals)

- ✓ Hadoop
- ✓ Spark
- ✓ Hive
- ✓ NoSQL(HBase, MongoDB)

Step -3 (Cloud)

Azure

- ✓ Blob Storage
- ✓ Data Factory
- ✓ Data Bricks

- ✓ Synapse Analytics

- ✓ SQL Database

- ✓ Stream Analytics

- ✓ Event Hubs, CosmosDB(optional)

AWS	**GCP**
✓ S3	BigQuery
✓ Athena	Cloud Storage
✓ Glue	DataFlow
✓ EMR	Dataproc
✓ Redshift	Pub/sub
✓ Lambda	Composer
✓ RDS Database	Data Fusion
✓ Kinesis	

Step -4

- ✓ Apache Kafka
- ✓ ETL Basics

Step -5

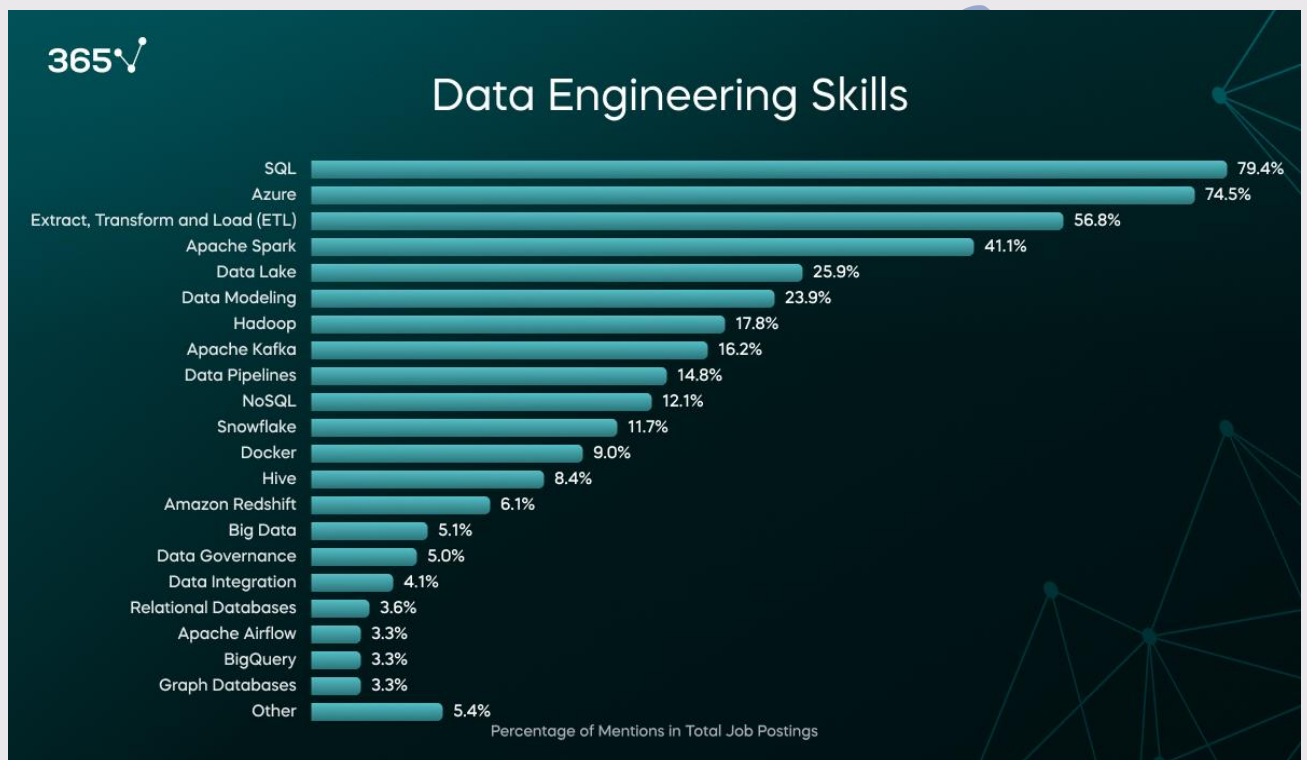- ✓ Apache Airflow

Step -6
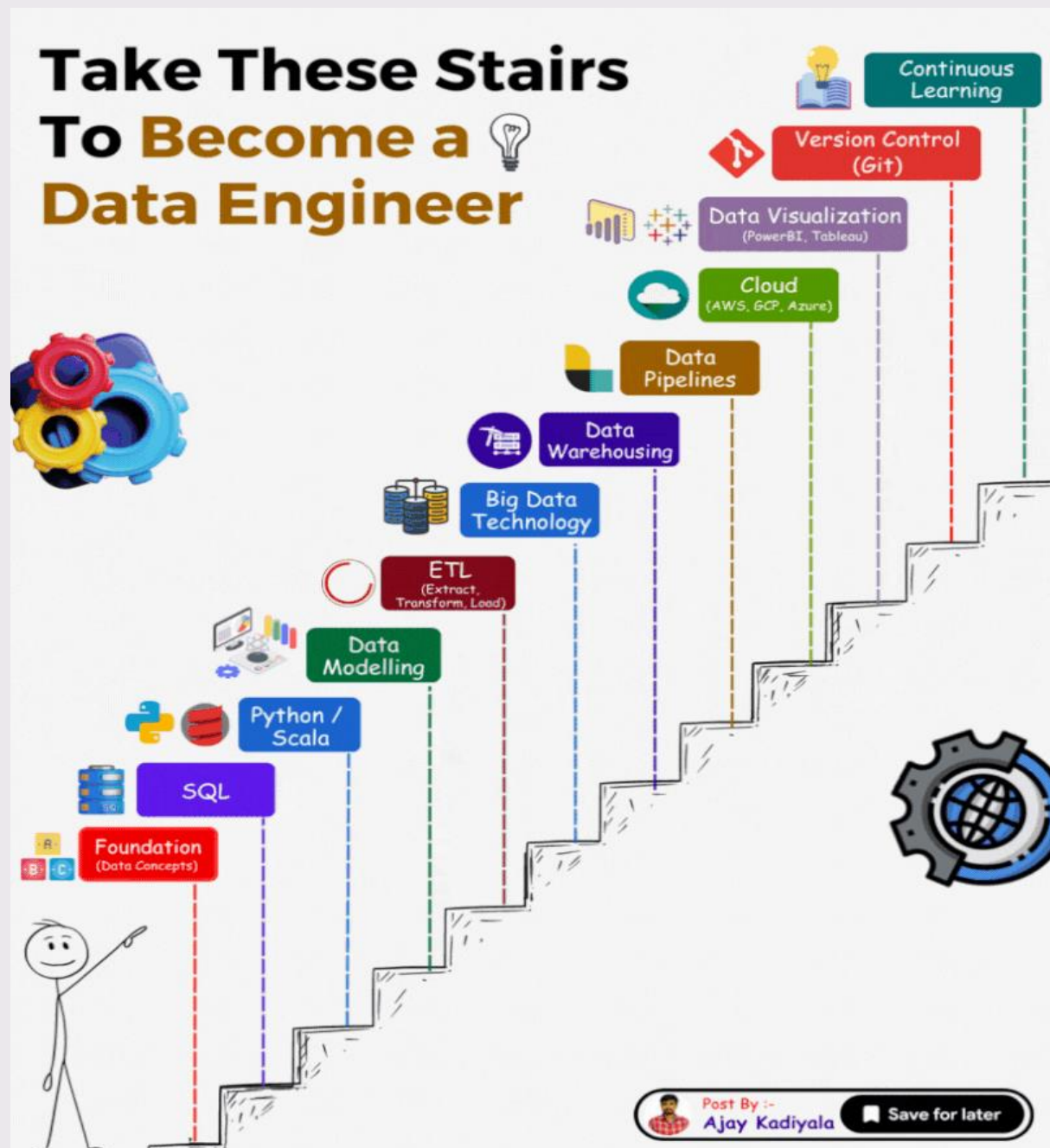
- ✓ CI/CD (GitHub, Docker, Kubernetes)

Step -7

- ✓ Data Modelling, Data Warehousing (star Schema, Snowflake Schema), Data Cleaning, Data profiling etc.

Step -8 (Soft Skills):

- ✓ Problem-solving and analytical mindset.

- ✓ Communication skills to collaborate with teams.

- ✓ Time management for handling complex tasks.

Take These Stairs To Become a Data Engineer

- Foundation (Data Concepts)
- SQL
- Python / Scala
- Data Modelling
- ETL (Extract, Transform, Load)
- Big Data Technology
- Data Warehousing
- Data Pipelines
- Cloud (AWS, GCP, Azure)
- Data Visualization (PowerBI, Tableau)
- Version Control (Git)
- Continuous Learning

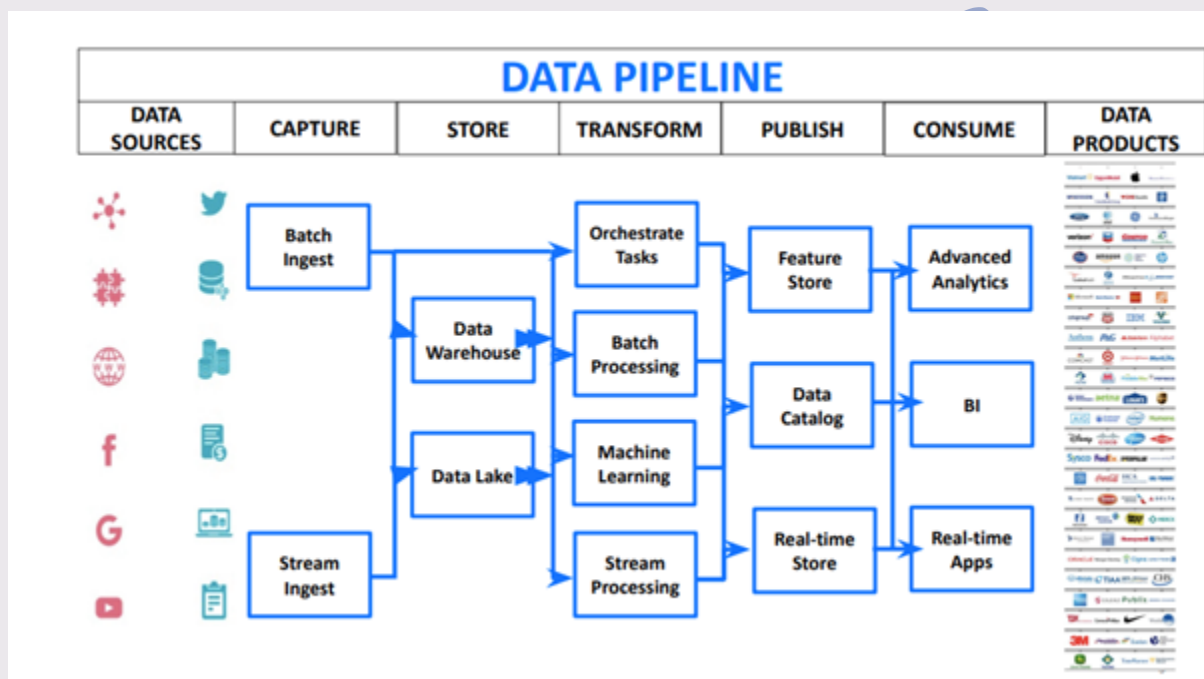Post By :- Ajay Kadiyala | Save for later

What Is Data Pipeline?

A **data pipeline** is a series of steps that moves data from one place to another, transforming it along the way so it's ready for analysis. Here's a simple breakdown:

1. **Data Ingestion:** Collect raw data from different sources like databases, APIs, or files.

2. **Data Processing:** Clean, transform, and organize the data to make it usable. This may include filtering, aggregating, or changing formats.

3. **Data Storage:** Save the processed data in a storage system, like a data warehouse or data lake, where it's easy to access.

4. **Data Analysis & Reporting:** Make the data available for analysis, visualizations, or machine learning to help drive insights and decisions.

Career Opportunities in Data Engineering:

- **Key Points:**
  - High demand for data engineers in industries like finance, healthcare, tech, and retail.
  - Average salary ranges from $90k to $150k depending on the role and region.
  - Job titles include: Data Engineer, Big Data Engineer, ETL Developer, and Cloud Data Engineer.

How much does a Data engineer make in India?

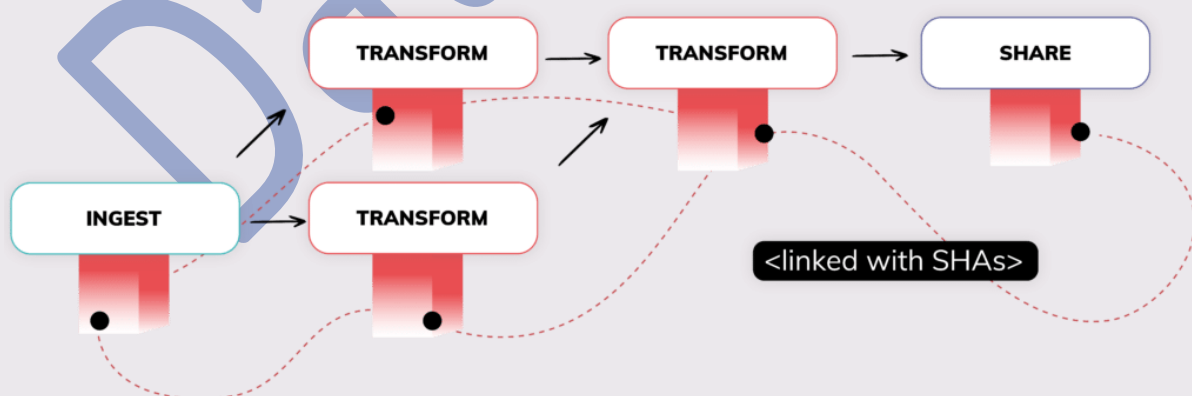### ₹ 1,500,000 / Annual

Based on 1573 salaries

The average **data engineer** salary in **India** is ₹ **1,500,000** per year or ₹ **769** per hour. Entry-level positions start at ₹ **1,000,000** per year, while most experienced workers make up to ₹ **2,578,833** per year.

Median
₹ 1,500,000

Low
₹ 1,000,000

High
₹ 2,578,833

Best Practices in Building Data Pipelines

- **Key Points**:
  - Automate data workflows where possible.
  - Ensure data quality checks at every stage.
  - Focus on scalability and optimizing performance.
  - Implement monitoring and logging to track pipeline health.

INGEST → TRANSFORM → TRANSFORM → TRANSFORM → TRANSFORM → SHARE

Learning Resources (FREE):

➤ **Evolution Of Data Engineering**

https://www.linkedin.com/posts/ajay026_dataengineering-trends-roadmap-activity-7246371339345821698-xnH8?Data

➤ **Data Engineering Roadmap:**

https://www.linkedin.com/posts/ajay026_dataengineering-comments-dataengineering-activity-7117717079914414080-7H0i?

➤ **Architecture of Data Engineering:**

https://www.linkedin.com/posts/ajay026_dataengineering-comments-dataengineering-activity-7123535097588285440-iPjg?

➤ **Top 10 AWS Services for data engineers:**

https://www.linkedin.com/posts/ajay026_dataengineering-github-dataengineering-activity-6993051655697043456-mWu_?

➤ **Data Engineering Jobs Growth Over the Years:**

https://www.linkedin.com/posts/ajay026_demand-dataengineering-demand-activity-7163868702520594432-Byng?

➤ **500+ Recent Interview Questions:**

https://www.linkedin.com/posts/ajay026_data-engineering-interview-questions-activity-7193950608251363328-pnDg?

➢ **9 Repositories to Build DE Portfolio:**

https://www.linkedin.com/posts/ajay026_dataengineering-python-pythonprogramminglanguage-activity-7039802545359265792-whzq?

➢ **Top Data Engineering Tools You should Learn:**

https://www.linkedin.com/posts/ajay026_dataengineering-dataengineering-tools-activity-7122906492244004864-ioGK?

➢ **Things you should learn for DE interviews**

https://www.linkedin.com/posts/ajay026_successful-dataengineering-visualization-activity-7084373249098207232-uwQ0?

➢ **How to build Data Pipelines**

https://www.linkedin.com/posts/ajay026_bigdata-data-dataengineering-activity-7041033560379572224-dzim?

➢ **Apache Spark Architecture**

https://www.linkedin.com/posts/ajay026_spark-dataengineering-machinelearning-activity-7039573038568259584-Re7d?

➢ **7 GCP Services to learn**

https://www.linkedin.com/posts/ajay026_dataengineering-gcpcloud-free-activity-6991241091190587392-dDcH?

➢ **Pyspark Roadmap**

https://www.linkedin.com/posts/ajay026_pyspark-interview-questions-activity-7197455988197580800-4Gms?

➢ **How Apache Spark Works internally**

https://www.linkedin.com/posts/ajay026_comments-dataengineering-data-activity-7116993328746827776-_-3c?

➢ **What is Big Data**

https://lnkd.in/gCutYTjV

➢ **Hadoop Architecture**

https://lnkd.in/g7cmKcdp

➢ **Hadoop Commands**

https://lnkd.in/gTMVKYUn

➢ **SQL Roadmap**

https://lnkd.in/gAGY-vX3

➢ **Learn Apache Hive**

https://lnkd.in/g_HbE6jJ

➢ **Learn Python**

https://lnkd.in/ggMZRfpf

➢ **Learn Scala Spark**

https://lnkd.in/gSzSiW6h

➢ **My Personal Interview Experience**

https://lnkd.in/gXRmkFtP

➢ **Recent SQL Interview Questions**

https://lnkd.in/gHB_bq48

➢ **Pyspark Interview Questions**

https://lnkd.in/ggpduwRv

➢ **Azure Data Factory Interview Questions**

https://lnkd.in/geaxscBs

➢ **Learn DSA**

https://lnkd.in/gz2PP8Z7

➢ **Complete Interview Guide**

https://lnkd.in/gy-nQ28W

➢ **Learn PowerBi**

https://lnkd.in/gV2Kn9M8

➢ **Learn Snowflake**

https://lnkd.in/gfmkej2B

➢ **SQL Projects**

https://lnkd.in/gPd_jkGU

https://www.linkedin.com/posts/ajay026_dataanalytics-dataanalysis-data-activity-7008638651542765568-tAjK?

➢ **Data Engineering Projects**

https://lnkd.in/g9hbFGUU

https://lnkd.in/g-WBDKVd

➢ **Bigdata Pipeline on AWS**

https://www.linkedin.com/posts/ajay026_bigdata-roadmap-dataengineering-activity-7039935426073169920-7T7e?

➢ **Leetcode SQL Questions**

https://www.linkedin.com/posts/ajay026_aceyoursqlinterviewsleetcode-activity-7158308454137270272-ZnYL?

- **Apache Airflow Roadmap**

  https://www.linkedin.com/posts/ajay026_apacheairflow-dataengineering-datapipelines-activity-7251434156314767360-g_L6?

- **Complete Azure Roadmap**

  https://www.linkedin.com/posts/ajay026_dataengineering-sql-python-activity-7024250825375961088-TsuY?

- **Walmart Interview Experience**

  https://www.linkedin.com/posts/ajay026_interviewpreparation-interviewquestions-dataengineering-activity-7167140997502525440-7Xvo?

- **Kafka Roadmap**

  https://www.linkedin.com/posts/ajay026_learn-to-work-with-apache-kafka-for-free-activity-7068835858753212416-s8GL?

- **Python Projects**

  https://www.linkedin.com/posts/ajay026_python-projects-with-source-code-aman-kharwal-activity-7024956726936313856-n0tV?

- **Python Data Analysis Projects**

  https://www.linkedin.com/posts/ajay026_dataanalytics-python-projects-activity-7010505103870738432-U5y5?

- **3 ETL Projects**

https://www.linkedin.com/posts/ajay026_dataanalytics-resume-sql-activity-7021000965864505345-vZ4J?

➢ **5 data engineering projects**

https://www.linkedin.com/posts/ajay026_dataengineering-projects-dataengineering-activity-7195996980672647170-T8lc?

➢ **Data Pipeline Project Lifecycle**

https://www.linkedin.com/posts/ajay026_data-piepline-life-cycle-activity-7061908174106214400-TtU3?

➢ **10 PowerBI Projects**

https://www.linkedin.com/posts/ajay026_dataanalytics-recommend-visualization-activity-7043461158481100800-0QWx?

➢ **Learn CI/CD here**

https://www.linkedin.com/posts/ajay026_integration-dataanalytics-data-activity-7045249030850809856-mR9K?

➢ **Learn Data Modelling, Data Warehousing.**

https://www.linkedin.com/posts/ajay026_dataengineering-datamodelling-datawarehouse-activity-7181498201638682624-_7rl?

➢ **Interviews Expectations for Data Engineers**

https://www.linkedin.com/posts/ajay026_dataengineering-interviews-dataengineering-activity-7132573673210716162-yWz-?

- **Amazon Interview Experience**

  https://www.linkedin.com/posts/ajay026_amazon-data-ai-activity-7071390880493961216-FQcY?

- **Uber Interview Experience**

  https://www.linkedin.com/posts/ajay026_data-uber-sql-activity-7251796734966673408-_Tjl?

- **Tesco Interview Experience**

  https://www.linkedin.com/posts/ajay026_systemdesign-interviewpreparation-interviews-activity-7167841897992892416-QplF?

- **Quantiphi Interview Experience**

  https://www.linkedin.com/posts/ajay026_quantiphi-collab-interviewpreparation-activity-7246723511463882752-dqCC?

- **Preparation Strategy for Data Engineers**

  https://www.linkedin.com/posts/ajay026_successful-company-dataengineering-activity-7082202684832845824-z5hl?

- **Learn Databricks**

  https://www.youtube.com/watch?v=bO7Xad1gOFQ&list=PLMWaZteqtEaKi4WAePWtCSQCfQpvBT2U1

Thank you

Q&A Session??